Gold

CBT Pollution Patch

By Liam Durant, ECE '20

Introduction

Team Gold's Senior Design Project was to create a sensor system for quantitative pollution detection. The system consists of two major subsystems. The first subsystem is an analog sensor. This sensor is a patch made out of cotton smart threads (Owyeung, Panzer, & Sonkusale). The patch is a grid of panels, with each panel made from a different cotton thread which has been dyed in a particular chemical. These chemical dyes are highly specific in that they change color when exposed to certain toxic industrial chemicals (TICS) (Lim, Feng, Kemling, Musto, & Suslick (2009). Moreover, this patch is an analog sensor that represents the desired pollutant data in the form of a color change relative to the unexposed control patch. The method of creation of the desired dyes used to make the patch have been well established by the team. The second subsystem is an android phone application that interprets the response from the analog sensor. This app is used to take a picture of the patch and then interpret the image using an image processing algorithm. The image processing algorithm detects and classifies the color changes of the panels and uses that information to determine what TICs the patch has been exposed to.

Problem Statement

In order to achieve the desired usage of the system, a method is needed to translate input from the analog sensor and mapping it to outputted results. This entails looking at how to map a set of normalized color data per panel to the overall projected TICs present. This process was broken down into a module that handles this translation. The creation of this module revolves around two components which greatly depend upon each other, a lookup table and signal constellation, and a Euclidean distance classification algorithm. These two concepts and their interaction will be outlined in this report.

Lookup Table and Signal Constellation

In order to be able to map real time data to a TIC and its corresponding concentration, data must exist that represents the known response for each TIC based on its concentration. This data is test data that our group hoped to obtain. This data generation process would include taking multiple identical patches to the lab and exposing them to different TICs and predefined concentrations, and recording the corresponding responses. And the filtered and formatted test data will then be stored in the lookup table. However, due to the circumstances that have arisen in early 2020, team gold was unable to gather this data but a solution was found. This solution included taking test data observed in another study (Lim, Feng, Kemling, Musto, & Suslick (2009)**.**

Data Generation

The two biggest questions our group had to address about the lookup table where: what type of data will be stored, and the size of the lookup table. The first question is quite complex; its resolution involved design decisions regarding the previous data filtering module in the application. Moreover, the input data to the lookup table is the output from the filtering module. We explored two main options: RGB values or wavelengths and their corresponding lumen intensity. RGB values were appealing as they are far less complex and more cost effective.

However, wavelengths were also appealing as it is likely they would provide more accurate data, as well as enabling highly specific and sensitive ranges of response wavelengths to be established for each pollutant. RGB values eventually had to be used since the lab became unavailable. Yet it is likely that RGB values would have been stored if the group had a choice. The second question revolved around the type of data chosen and the necessary amount to be stored per panel. The answer to this question would have been entirely answered by testing. Moreover, the data gathering process alone would have answered both of these questions as well as a number of uncertainties we had regarding the usage of the patches. Mainly, how linear is the color response of the patch to a TIC with respect to concentration present. It will also answer a major question regarding the differentiability of each TIC based on color response.

The lookup table was created by interpreting data collected in a different study. This study performed tests on 20 TICs, and mapped the color change of 5 dyes which we aimed to use on our patch. An example of one of the 20 entries is outlined in the figure below:


```
{
    "chemical": "Arsine 1",
    "R11": "-0.71",
    "G11": "-1.20",
    "B11": "1.29",
    "R23": "-0.39",
    "G23": "-0.06",
    "B23": "-0.16",
    "R24": "-0.89",
    "G24": "-0.49",
    "B24": "-0.69",
    "R25": "-0.84",
    "G25": "-0.11",
    "B25": "0.27",
    "R29": "-0.27",
    "G29": "-0.30",
    "B29": "-0.59"
},
```


Figure 1: Example lookup table entry

Signal Constellation

The lookup table was then used to construct a N-dimension signal constellation. Note I said we used the data type as RGB, so N = 3. The signal constellation will act as the map between the lookup table and the filtered input data (Porath & Aulin (2003)). The constellation puts these pieces of data all in the same space to be compared against. Our projected constellation is 3-dimensional, with one dimension for red, green, and blue.

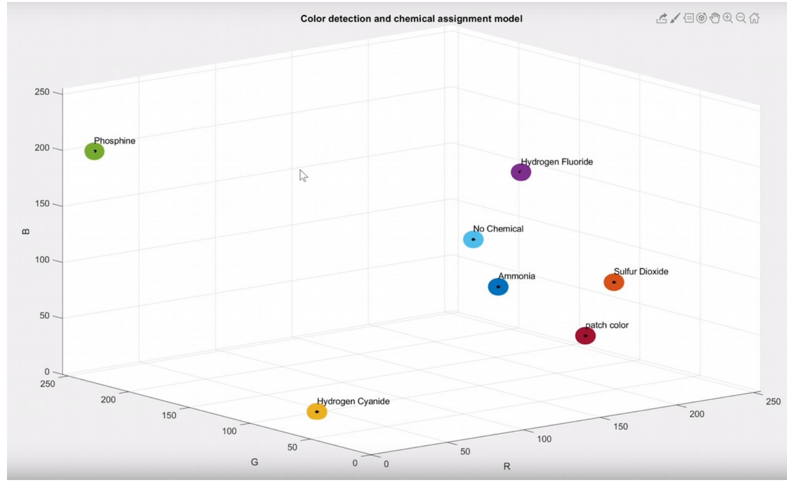

Figure 2: 3D signal constellation example in Matlab

Figure 2 outlines the visual representation of the 3D-signal constellation for a made up panel on the patch. Notice there are 5 example TICs present, one control, and the input “patch color” point. A number of distinctions need to be made about this model. The first is that a constellation exists for each panel. So the “patch color” point would actually refer to “panel color”. Next this is an arbitrary made up panel, that responds to the arbitrary TICs in the manner that they are presented. Lastly, this model visualizes the TIC concentration in the form of a spherical cloud: where this most likely would not have been the case after gathering test data. For example, if testing determined a linear color response per TIC (as is expected of the dyes (Owyeung, Panzer, & Sonkusale)), then the color points would be modeled with a cone or cylinder in space. With one end representing the bottom of the range of TIC concentration and the other end representing the top of the range, giving that defined space as the presence of a chemical. The width of the cone or cylinder in this case would be entirely separate of the direction of the concentration vector and completely dependent on our models of certainty of TIC presence. This model is essential as it allows us to draw conclusions about TIC concentrations by using a Euclidean distance classification algorithm.

Euclidean Distance Classification Algorithm

With our signal constellation established, a euclidean distance classification algorithm can be used to draw conclusions about TIC concentrations. On the surface this algorithm is a very simple calculation. The algorithm will work as follows: for each panel signal constellation, calculate the euclidean distance between the "color panel" data point and each subsequent TIC point in the constellation (Porath & Aulin (2003)). So from our example from figure 1, the algorithm would yield five scalar values, one associated with each TIC. This algorithm can be done in O(n) time per panel, where n is the amount of TICs present in the constellation. Now with the distances calculated for each panel, a meaningful comparison needs to be made between the panels and their subsequent distances to figure out the concentrations of present TIC. This was a design challenge that would be solved via experimental results.

Due to the restrictions of present test data, major adjustments had to be made with regard to this algorithm. Moreover, the euclidean distance calculation was adapted to be performed per chemical rather than per panel. This yielded a very simple but effective classification algorithm: calculate the euclidean distance of color change over the 5 panels per each TIC and choose the TIC (or control) with the smallest distance as the present TIC. Obviously a large amount of error would be associated with this algorithm. A threshold distance would need to be set per chemical, which more experimental data could provide.

Implementing a classification algorithm in this way presents a large issue revolving around imperfect information. Furthermore, differentiating cumulative distances per panel with more than one TIC response being present is a major issue. And with our limited data, we can only set our algorithm to output the TIC with the highest certainty of being present.

Conclusion

The problem was to implement a method that could translate digital values from an analog sensor to a decision regarding the presence of TICs. There are many different approaches to generating a classifier capable of performing this operation. However, a simple lookup table and euclidean distance classification algorithm performed the necessary action. Its selection was made based on its simplicity, performance, and cost effectiveness both in computation time and memory management. With the proper experimental data present, this algorithm would have provided the team with the desired usage of a TIC output classifier.

References

1. Lim, S. H., Feng, L., Kemling, J. W., Musto, C. J., & Suslick, K. S. (2009). An optoelectronic nose for the detection of toxic gases. *Nature Chemistry, 1*(7), 562–567. https://doi.org/10.1038/nchem.360

2. Owyeung, R. E., Panzer, M. J., & Sonkusale, S. R. (2019). Colorimetric Gas Sensing Washable Threads for Smart Textiles. *Scientific Reports*, *9*(1), 5607. https://doi.org/10.1038/s41598-019-42054-8

3. Porath, J.-E., & Aulin, T. (2003). Design of multidimensional signal constellations. *IEE Proceedings - Communications*, *150*(5), 317-. https://doi.org/10.1049/ip-com:20030652